ON k-p-INFIX CODES (Algebraic Systems, Formal Languages and Computations)

数理解析研究所講究録
1166 巻 2000 年 152-160

ON k-p-INFIX CODES

Dongyang Long and **Weijia Jia**
Department of Computer Science, City University of Hong Kong,
83 Tat Chee Avenue, Kowloon, Hong Kong, China
Email:{*dylong, wjia*}*@cs.cityu.edu.hk*

ABSTRACT: *In this paper we introduce the concepts of k-p-infix codes, n-k-ps-infix languages, n-k-infix-outfix codes, and n-k-prefix-suffix languages, which are natural generalizations of our previous work on k-prefix codes, k-infix codes and so on. We obtain several properties of k-p-infix codes and semaphore codes. The relations and hierarchies of k-p-infix codes, n-k-ps-infix languages, n-k-infix-outfix codes, and n-k-prefix-suffix languages, and their operations of these classes of languages are also investigated.*
KEY WORDS: *code, shuffle code, combinatorics on words, product of languages.*

1. Introduction

Codes and languages derived from or related to codes have an important role in the study of the combinatorics of words [6]. Many classes of codes can be obtained as the classes of antichains with respect to certain partial orders on free monoids [2-5, 9-13]. In particular, various kinds of classes of codes defined by insertion properties and their corresponding hierarchy properties were given [6]. There are much work related to the topic such as n-codes [2-3], n-prefix-suffix languages [5], n-infix-outfix codes [9-10], and k-shuffle codes [6-8, 15]. Especially, as pointed out in recent survey paper [6], these variations on insertion properties are more than just generations for all kinds of different names in earlier publications, but have concrete implications for the error detection capabilities of such codes. Hence they are quite interesting also in a broader sense. The ideal of studying n-codes and n-k-languages is very natural, a main motivation of this paper aims to extend the authors previous work on k-prefix codes, k-infix codes and so on.

We first introduce the necessary concepts and notations. For additional details and definitions, see the references, in particular [1], [5], [6], and [14].

Let A be a finite alphabet and $L \subseteq A^*$ be a language. Denote $A^+ = A^* - \{1\}$ where 1 is the empty word over A. For a language L one associates with its syntactic monoid syn(L) = A^* / P_L where

$$x \equiv y(P_L) \Leftrightarrow (\forall u.v \in A^*) uxv \in L \leftrightarrow uyv \in L$$

By $[w]$ we denote the P_L -class of the word w, i.e. $[w] = \{ x \in A^* | x \equiv w(P_L) \}$. For every $w \in A^*$, we denote by $|w|$ the length of w.

A language $L \subseteq A^*$ is said to be a code over A if the submonoid L^* of A^* generated by L is freely generated by L. If P is any property of languages, we call a code C a P-code if C possesses the property P. If C is a P-code and, for every u ($\notin C$) $\in A^*$, $C \cup \{u\}$ is not a P-code, then C is said to be a maximal P-code.

Definition 1. [6-8] *Let A be an alphabet and k be a given positive integer, A language* $C \subseteq A^*$ *is said to be*

(a) *a k-prefix code if for all* $x_1,\ldots,x_k, y_1,\ldots,y_k \in A^*, x_1\ldots x_k \in C$ *and* $x_1y_1x_2\ldots x_ky_k \in C$ *together imply* $y_1\ldots y_k = 1$;

(b) *a k-suffix code if for all* $x_1,\ldots,x_k, y_1,\ldots,y_k \in A^*, x_1\ldots x_k \in C$ *and* $y_1x_1\ldots y_kx_k \in C$ *together imply* $y_1\ldots y_k = 1$;

(c) *a k-infix code if for all* $x_1,\ldots,x_k, y_0,\ldots,y_k \in A^*, x_1\ldots x_k \in C$ *and* $y_0x_1y_1\ldots x_ky_k \in C$ *together imply* $y_0y_1\ldots y_k = 1$;

(d) *a k-outfix code if for all* $x_0,\ldots,x_k, y_1,\ldots,y_k \in A^*, x_0x_1\ldots x_k \in C$ *and* $x_0y_1x_1\ldots y_kx_k \in C$ *together imply* $y_1\ldots y_k = 1$;

(e) *a hypercode if for any natural nnumber n and all* $x_1,\ldots,x_k, y_0,\ldots,y_k \in A^*, x_1\ldots x_k \in C$ *and* $y_0x_1\ldots x_ky_k \in C$ *together imply* $y_0\ldots y_k = 1$;

(f) *a full uniform code if there exists some integer* $m \geq 0$ *such that* $C = A^m$.

By $P_k(A)$, $S_k(A)$, $I_k(A)$, $O_k(A)$, $H(A)$ and $FUF(A)$ we denote the classes of k-prefix codes, k-suffix codes, k-infix codes, k-outfix codes, hypercodes and full uniform codes over A, respectively [7-8] . In particular, $P(A) = P_1(A)$, $S(A) = S_1(A)$, $I(A) = I_1(A)$, $O(A) = O_1(A)$ are the classes of prefix, suffix, infix, and outfix codes, respectively.

Note that k-prefix codes, k-suffix codes, k-infix codes, and k-outfix codes are also called prefix-shuffle, suffix-shuffle, infix-shuffle, and outfix-shuffle codes of index k, respectively [6] [15]. And corresponding classes of codes are denoted by L_{Pk} ($=P_k(A)$), L_{Sk}($=S_k(A)$), L_{Ik}($= I_k(A)$), and L_{Ok}($=O_k(A)$). In [6], by L_h and L_u denote hypercodes and uniform codes over A. Relations between these codes can be referred to Fig. 6.1 and 7.1 in Chapter 8 of [6].

Definition 2. [6-8] *Let A be an alphabet. A language* $C \subseteq A^*$ *is said to be*

(a) *a bifix (or biprefix) code if C is both a prefix and a suffix code;*

(b) *reflective if for all* $u, v \in C$ *imply* $vu \in C$;

(c) *a p-infix code if for all* $x, u, y \in A^*, xuy \in C$ *and* $u \in C$ *together imply* $y = 1$;

(d) *a s-infix code if for all* $x, u, y \in A^*, xuy \in C$ *and* $u \in C$ *together imply* $x = 1$;

(e) *a right semaphore code if C is a prefix code satisfying* $A^*C \subseteq CA^*$;

(f) *a left semaphore code if C is a suffix code satisfying* $CA^* \subseteq A^*C$.

By $B(A)$, $RE(A)$, $PI(A)=PI_1(A)$, $SI(A) = SI_1(A)$, $RSP(A)$ and $LSP(A)$ denote the classes of bifix, reflective, p-infix, s-infix, right semaphore and left semaphore codes over A, respectively.

Note that, in [6], by $L_b(= B(A))$, $L_{refl}(= RE(A))$, $L_{pi}(= PI(A))$, $L_{si}(= SI(A))$, $L_{rsema}(= RSP(A))$ and $L_{lsema}(= LSP(A))$ denote the classes of bifix, reflective, p-infix, s-infix, right semaphore and left semaphore codes over A, respectively, Relations between the above codes can be referred to Fig. 7.2 in Chapter 8 of [6].

The paper is organized as follows: After introduction section, we introduce the classes of k-p-infix and k-s-infix codes. The relations and hierarchies of k-p-infix, k-s-infix, right semaphore and left semaphore codes are given in Section 2. In Section 3, the hierarchy of n-k-ps-infix codes is obtained, which is a natural generalization of k-p-infix and k-s-infix codes. In Section 4, we investigate n-k-infix-outfix and n-k-prefix-suffix languages. Meanwhile, their hierarchies and product properties of two classes of languages are also discussed.

2. *k*-p-Infix Codes

Definition 3. *A languages $L \subseteq A^*$ is said to be a k-p-infix (k-s-infix) code if for all $x_1,\ldots,x_k, y_1,\ldots,y_k, y \in A^*, x_1\ldots x_k \in C$ and $y_1x_1y_2\ldots y_kx_ky \in C (yx_1y_1\ldots x_ky_k \in C)$ together imply $y=1$.*

From Definition 3 it easily follows that a (k+1)-p-infix code must be a k-infix code. By $PI_k(A)$ ($SI_k(A)$) we denote the class of k-p-infix (k-s-infix) codes over A. Therefore, we have

Theorem 1. $PI_1(A) \supset PI_2(A) \supset PI_3(A) \supset \ldots \supset PI_k(A) \supset PI_{k+1}(A) \supset \ldots$.

Proof: Since $PI_k(A) \supseteq PI_{k+1}(A)$, it suffices to show that there exists $C \in PI_k(A)$ such that $C \notin PI_{k+1}(A)$. Let $A = \{a, b\}$, $C = \{a^{k+1}, (ab)^{k+1}\}$. We can easily verify that $C \in PI_k(A)$ but $C \notin PI_{k+1}(A)$.

Theorem 2. *The $PI_k(A)$ is closed under product, that is the $PI_k(A)$ forms a monoid. Conversely if XY is a k-p-infix code then both X and Y need not be k-p-infix codes.*

Proof: Let $X, Y \in PI_k(A)$. If, for all $u_1,\ldots,u_k, v_1,\ldots,v_k \in XY$ and $v_1u_1v_2\ldots v_ku_kv \in XY$, then there exist $x_1, x_2 \in X$ and $y_1, y_2 \in Y$ such that $u_1\ldots u_k = x_1y_1$ and $v_1u_1v_2\ldots v_ku_kv = x_2y_2$. Let $u_1\ldots u_{i-1}u_i' = x_1$ and $u_i''u_{i+1}\ldots u_k = y_1$ with $u_i = u_i'u_i''$. If $|x_2| > |v_1u_1v_2u_2\ldots u_{i-1}v_iu_i'|$, then $x_2 = v_1u_1v_2u_2\ldots u_{i-1}v_iu_i'w$ with $w \in A^+$. Since $X \in PI_k(A)$ and $i \le k$, thus $w = 1$, a contradiction with $w \in A^+$! Therefore $|x_2| \le |v_1u_1v_2u_2\ldots u_{i-1}v_iu_i'|$ and $y_2 = wu_i''v_{i+1}u_{i+1}\ldots u_kv$ for some $w \in A^*$. But $y_1 = u_i''u_{i+1}\ldots u_k$ and $Y \in PI_k(A)$, we have $v = 1$. This shows that $XY \in PI_k(A)$. That is, the $PI_k(A)$ is closed under product and consequently forms a monoid.

Conversely, let $A = \{a, b\}$, $XY = \{a^{k+1}, (ba)^{k+1}\}$, then we can directly verify that XY is a k-p-infix code. When we take $X = \{a^k, (ba)^k b\}$ and $Y = \{a\}$, it is easy to see that X is not a k-p-infix code but Y is a k-p-infix code. Clearly, when we take $X = \{1\}$ and $Y = XY$, then X and Y are k-p-infix codes.

From definitions and Theorem 3 in [7], it easily follows that

Theorem 3. *Let $C \in PI_k(A)$. Then*
(1) *C is an infix code if and only if C is a suffix code.*
(2) *C is a full unform code, that is $C = A^m$ for some m, if and only if C is a maximal suffix code.*

By Proposition 5.3 in Chapter 2 of [1] and Theorem 1, we immediately obtain the following theorem.

Theorem 4. *Let $C \in PI_k(A)$. Then C is a right semaphore code if and only if C is a maximal prefix code.*

Theorem 5. *Any k-p-infix code is thin.*

Proof: By Theorem 1, we see that the class of 1-p-infix codes contains the classes of k-p-infix codes for $k \geq 2$. Since a 1-p-infix code is thin, by definition, a k-p-infix code is thin.

Corollary1. *Let $C \in PI_k(A)$. Then C is a right semaphore code if and only if C is a maximal code.*

By duality, we have

Theorem 6. (1) $SI_1(A) \supset SI_2(A) \supset SI_3(A) \supset \ldots \supset SI_k(A) \supset SI_{k+1}(A) \supset \ldots$
(2) *The $SI_k(A)$ is closed under product.*
(3) *Let $C \in SI_k(A)$. Then C is an infix code if and only if C is a prefix code.*
(4) *Let $C \in SI_k(A)$. Then C is a full uniform code, that is $C = A^m$ for some m, if and only if C is a maximal prefix code.*
(5) *Let $C \in SI_k(A)$. Then C is a right semaphore code if and only if C is a maximal suffix code.*
(6) *Any k-s-infix code is thin.*
(7) *Let $C \in SI_k(A)$. Then C is a right semaphore code if and only if C is a maximal code.*

On finite k-p-infix codes, we have

Theorem 7. *Let X be a finite k-p-infix code. Then* $X' = X_1 \cup X_2A^{-1}$ *is a k-p-infix code, where* $X_1 = X - X_2$, $X_2 = \{x \in X | (\forall x \in X) |x'| \leq |x|\}$.

Proof: Arguing by contradiction, we assume that there exist $u_1,...,u_k,v_1,...,v_k \in A^*$, $v \in A^+$, such that $v_1...v_k \in X'$ and $v_1u_1...v_ku_kv \in X'$. (i) if $v_1...v_k$, $v_1u_1...v_ku_kv \in X_1$, since $X \in PI_k(A), v = 1$ which is impossible. (ii) if $v_1...v_k$, $v_1u_1...v_ku_kv \in X_2A^{-1}$ then there exist $a, b \in A$ such that $v_1...v_k a$, $v_1u_1...v_ku_kvb \in X_2$, contradicting with the choice of X_2. (iii) if $v_1...v_k \in X_1$ and $v_1u_1...v_ku_kv \in X_2A^{-1}$, then $v_1...v_k \in X_1$ and $v_1u_1...v_ku_kva \in X_2$, for some $a \in A$. This is a contradiction with X being k-p-infix code. Clearly, if $v_1...v_k \in X_2A^{-1}$ then $v_1u_1...v_ku_kv \notin X_1$. Thus we show that X' is a k-p-infix code.

By definitions, we can easily following Lemma 1

Lemma 1. *Let* $X \subseteq A^*$. *Then X is a maximal 1-p-infix code if and only if*

$$A^* = X \cup A^*XA^+ \cup (A^*)^{-1}X(A^+)^{-1}$$

We will give anther characterization of right semaphore codes which is different from that in [1].

Theorem 8. *Let* $X \subseteq A^+$. *Then X is a right semaphore code if and only if X is a maximal 1-p-infix code.*

Proof: We first show that if X is a maximal 1-p-infix code then X must be a right semaphore code. Let $S = X - A^+X$. Clearly S is a nonempty subset of X. To prove that X is a right semaphore code, let us show that $X = A^*S - A^*SA^+$.
By definition of $S, X \subseteq A^*S$. Since $S \subseteq X$ and X is 1-p-infix, $X \cap A^*SA^+ = \varnothing$. This shows that $X \subseteq A^*S - A^*SA^+$. Assume that there exists a word y in $(A^*S - A^*SA^+) - X$. By hypothesis, $\{y\} \cup X$ is not 1-p-infix. Either $y = uxv$ or $x = uyv$ with $x \in X, u \in A^*, v \in A^+$. In the first case, since $x \in A^*S$, it follows that $y \in A^*SA^+$ which is impossible. In the second case, $y \in A^*S$ means that $x \in A^*SA^+$, a contradiction with $X \subseteq A^*S - A^*SA^+$. Hence $X = A^*S - A^*SA^+$. This shows that X is a right semaphore.
Conversely, assume that X is a right semaphore code, then it is 1-p-infix. Suppose that X is not a maximal 1-p-infix code, there exists $y \in A^* - X$ such that $\{y\} \cup X$ is a 1-p-infix. By the definition of a 1-p-infix code, $\{y\} \cup X$ is a prefix code. But X is a right semaphore code, and consequently X is a maximal prefix, a contradiction with $\{y\} \cup X$ being a prefix code. That is, X is a maximal 1-p-infix code.

Remark 1. By Theorem 8, clearly, a maximal 1-p-infix code must be a maximal prefix code. Conversely, in general, a maximal prefix code need not be 1-p-infix code.

Corollary 2. *Let* $X \subseteq A^+$ *. Then X is a left semaphore code if and only if X is a maximal 1-s-infix code.*

Remark 2. Let $X, Y \subseteq A^*$ be maximal k-p-infix codes for $k \geq 2$. Then XY need not be a maximal k-p-infix code.

Remark 3. By definition, a right semaphore code must be a 1-p-infix code. But a 1-p-infix code is not necessarily a right semaphore code.

From Remarks 2 to 3, it seems to see that there are many differences between 1-p-infix codes and k-p-infix codes for $k \geq 2$, although we have Theorem 1. Therefore, the study of relations between 1-p-infix codes, k-p-infix codes for $k \geq 2$ and semaphore codes will be very interesting.

3. n-k-ps-Infix Languages

Similar to n-prefix-suffix languages [5], we define

Definition 4. *A language* $X \subseteq A^*$ *is said to be a n-k-ps-infix code, if every subset of X at most n elements is a k-p-infix code or a k-s-infix code.*

By k-$PSI_n(A)$ we denote the class of n-k-ps-infix codes. We have

Theorem 9.
$k-PSI_2(A) \supset k-PSI_3(A) \supset k-PSI_4(A) = k-PSI_5(A) = ... = PI_k(A) \cup SI_k(A).$

Theorem 10.
(1) *k-$PSI_2(A)$, k-$PSI_3(A)$, and k-$PSI_4(A)$ = k-$PSI_5(A)$ are not closed under product.*
(2) *Both k-$PSI_2(A)$ and k-$PSI_3(A)$ need not be codes.*

By Theorems 3 and 5, we can directly follow that

Corollary 3. *Let* $X \in k$*-$PSI_4(A)$* $= PI_k(A) \cup SI_k(A)$*. Then X is an infix code if and only if X is a biprefix code.*

Corollary 4. *Any 4-k-ps-infix code is thin.*

Corollary 5. *Let* $X \in k$*-$PSI_4(A)$* $= PI_k(A) \cup SI_k(A)$*. Then X is a full uniform code, that is X =* A^m *for some m, if and only if X is a maximal biprefix(or bifix) code.*

Remark 4. Fig. 1 illustrates the relations between n-k-ps-infix codes

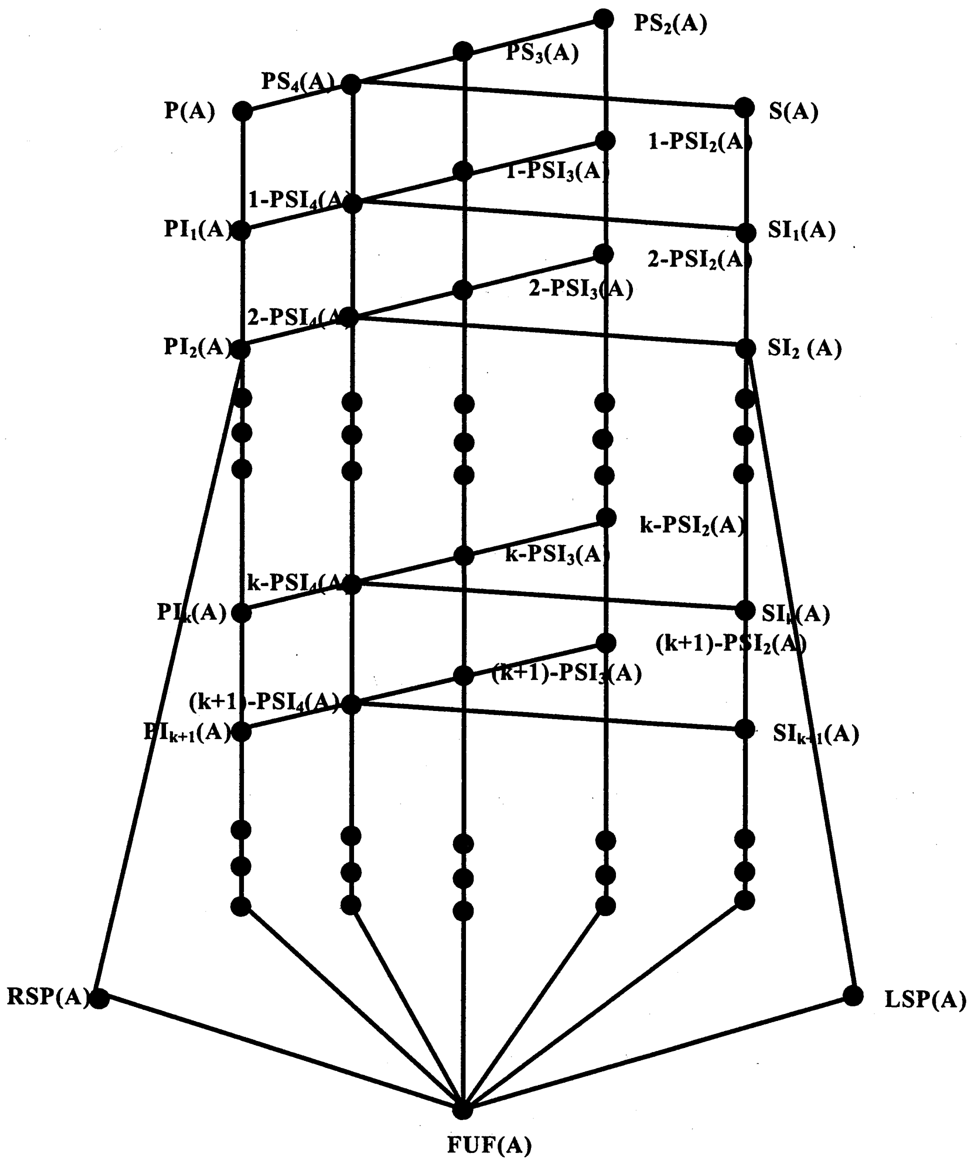

Fig. 1 Relations between n-k-ps-infix codes

4. n-k-Infix-Outfix and n-k-Prefix-Suffix Languages

Definition 5. (1) *A language $X \subseteq A^*$ is said to be a n-k-infix-outfix code if every subset of X at most n elements is a k-infix code or a k-outfix code.*
(2) *A language $X \subseteq A^*$ is said to be a n-k-prefix-suffix code if every subset of X at most n elements is a k-prefix code or a k-suffix code.*

By $k\text{-}IO_n(A)$ ($k\text{-}PS_n(A)$) we denote the class of the n-k-infix-outfix (n-k-prefix-suffix) codes over A. In particular, $1\text{-}IO_n(A)$ ($1\text{-}PS_n(A)$) is the class of n-infix-outfix (n-prefix-suffix) codes [5, 9].

From the definitions we easily follows

Theorem 11. (1) $k-IO_2(A) \supset k-IO_3(A) \supset k-IO_4(A) = k-IO_5(A) = I_k(A) \cup O_k(A)$.
(2) $k-PS_2(A) \supset k-PS_3(A) \supset k-PS_4(A) = k-PS_5(A) = P_k(A) \cup S_k(A)$.
(3) *k-$IO_2(A)$, k-$IO_3(A)$, and k-$IO_4(A)$ are closed under product. Conversely, if $XY \in k$-$IO_2(A)$ (k-$IO_3(A)$ and k-$IO_4(A)$) then X and Y need not be in k-$IO_2(A)$ (k-$IO_3(A)$ and k-$IO_4(A)$).*
(4) *In general, k-$PS_3(A)$ and k-$PS_4(A)$ = k-$PS_5(A)$ = $P_k(A) \cup S_k(A)$ are not closed under product.*

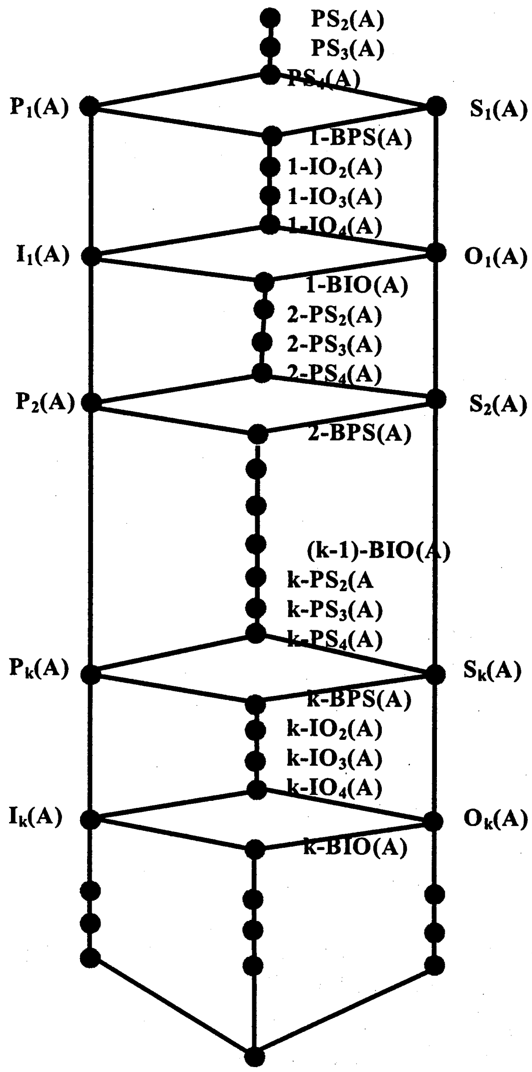

Fig. 2 Relations between n-k-infix-outfix codes and n-k-prefix-suffix codes

Remark 5. On $k\text{-}PS_2(A)$, there is a complex situation. We can easily show that the class of $1\text{-}PS_2(A)$ is not closed under product. However, on $k\text{-}PS_2(A)$ for $k \geq 2$, we have neither obtained an example which shows that $k\text{-}PS_2(A)$ for $k \geq 2$ is not closed under product, and nor proved that $k\text{-}PS_2(A)$ for $k \geq 2$ is closed under product.

Remark 6. Fig.2 illustrates the relations between n-k-infix-outfix codes and n-k-prefix-suffix codes. Especially, relations among Fig.1, Fig.2, and some classes of languages derived from codes can be referred to Fig. 7.2 and Table 8.1 in Chapter 8 of [6].

Acknowledgement

This work was partially sponsored by the City University of Hong Kong under grants 7000853, 7000765, 7100065 and RGC Hong Kong grant under 9040352. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official polices or endorsements, either express or implied, of the UGC Hong Kong or the City University of Hong Kong.

References

[1] J. Berstel, D. Perrin, *Theory of Codes*(Academic Press, New York, 1985).

[2] M. Ito, H. Jürgensen, H.J. Shyr, G.Thierrin, *Anti-commutative Languages and n-codes,* Discrete Appl. Math. **24**(1989),187-196.

[3] M. Ito, H. Jürgensen, H.J. Shyr, G. Thierrin, *Languages Whose n-Element Subsets Are Codes*, Theoretical Computer Science, **94**(1992), 325-349.

[4] M. Ito, H. Jürgensen, H.J. Shyr, G. Thierrin, *Outfix and Infix Codes and Related Classes of Languages*, J. Comput. System Sci. **43**(1991), 484-508.

[5] M. Ito, H. Jürgensen, H.J. Shyr, G. Thierrin, *n-Prefix-Suffix Languages*, International J of Computer Mathematics, **30**(1989), 37-56.

[6] H. Jürgensen, S. Koustantinidis, *Codes,* in: G.Rozenberg, A. Salomaa ed., Handbook of Formal Languages (Springer-Verlag, Berlin,1997) 511-607.

[7] D.Y. Long, *On the Structure of Some Group Codes*, Semigroup Forum **45**(1992), 38-44.

[8] D.Y. Long, *On Two Infinite Hierarchies of Prefix Codes*, in: K.P.Shum ed., Proc. Conf. on Ordered Structures and Algebra of Computer Languages, (World Scientific, Hong Kong, 1993) 81-90.

[9] D.Y. Long, *N-infix-outfix Codes*, Chinese Science Bulletin **39**(1994),584-586.(in Chinese)

[10] D.Y. Long, J. Ma, D.N. Zhou, *Structure of 3-infix-outfix Maximal Codes*, Theoretical Computer Science **188**(1997) 231-240.

[11] D.Y. Long, *On Group Codes*, Theoretical Computer Science **163**(1996) 259-267.

[12] D.Y. Long, *k-outfix Codes*, Chinese Ann. Math. Ser. **A** **10**(1989), 94-99(in Chinese)

[13] D.Y. Long, k-prefix Codes and k-infix Codes, Acta Math. Sinica, **33**(1990), 414-421(in Chinese)

[14] H.J. Shyr, *Free Monoids and Languages*, (Hon Min Book Co. Taichung, Taiwan, 1991).

[15] G. Thierrin and S.S. Yu, *Relations and Codes*, Journal of Information & Optimization Sciences **12**(1991)3, 441-449.